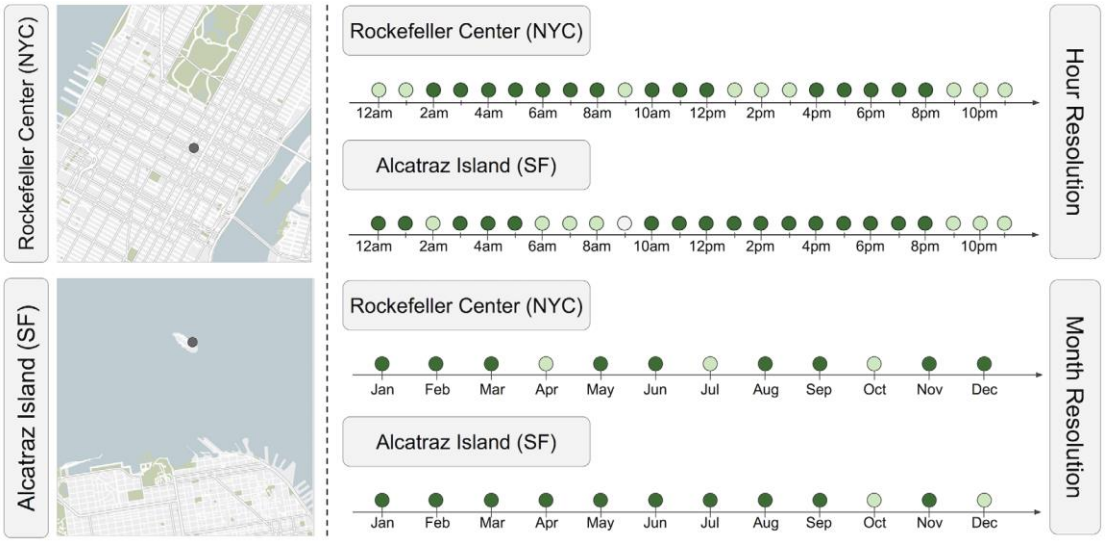
Weekly report

1 Done

1.1 Reading

1.1.1 Urban Pulse: Capturing the Rhythm of Cities (Fabio Miranda)



The idea of this paper is quite similar with Teacher Wu. They employ Gaussian weighted sum to define density function as well:

$$f(p) = \sum_{x_i \in N(p)} e^{\frac{-d(p,x_i)^2}{\varepsilon^2}}$$

What they do better than us is that they care more details listed as follow:

- They have definitions of terms like significant beats, maxima beats and critical points.
- Their interface is well-designed.
- “Urban Pulse” is an appropriate name to generalize the work and attract readers.
- Their dataset has the ability to provide analysis about multiple temporal resolutions including month of year, day of week and hour of day, which is of great significance.

1.1.2 On Syntactic Anonymity and Differential Privacy (Chris Clifton)

This paper provides a comprehensive summary of privacy-preserving approach like syntactic anonymity (privacy-preserving data publishing (PPDP)) and differential privacy (privacy-preserving data mining (PPDM)) and their limitations.

- Approach mentioned: PPDP including k-anonymity, l-diversity, t-closeness, β-likeness and p-sensitivity; PPDM like ε-differential privacy.
- Limitation of syntactic anonymity:

- May not provide sufficient privacy. Composition attacks, minimality attacks and the deFinetti attack are mentioned.
- The curse of dimensionality: high-dimensional data would significantly degrade the data quality.

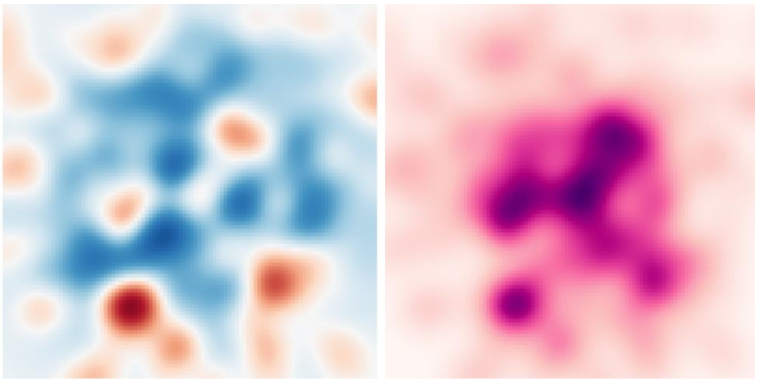
● Limitation of differential privacy:

- It may affect the correlations so that independence assumption is necessary.
- Non-compact uncertainty: sometimes bounds of original data is useful.
- Parameter setting is hard when it comes to problem about geographic units and something else.

1.1.3 Surprise! Bayesian Weighting for De-Biasing Thematic Maps (Michael Correll)

The objects that representing a large change to user beliefs entails high Bayesian surprise. This paper calculates the value of "surprise" based on Bayesian weighting and KL-divergence (shown as below).

$$KL(P(\mathscr{M}|D)||P(\mathscr{M})) = \sum_{i=1}^{|\mathscr{M}|} P(M_i|D) \log \frac{P(M_i|D)}{P(M_i)}$$



Besides, they use different color encoding for Surprise and KDE density. As for Surprise, they employ bright reds and blues for the extrema of the scales.

1.1.4 Privacy-Preserving Deep Learning (Reza Shokri)

This paper provides a practical system for collaborative deep learning that offers an attractive tradeoff between utility and privacy. Deep learning in this paper is facilized to control the parameter in differential privacy. However, the deep learning part is confused to me, so that next week I will start from the fundamental knowledge.

1.1.5 Reverse-Engineering Visualization (EuroVis Submission)

The major contributions of this paper include a text analysis pipeline (identifies text elements in a chart image, determines their bounding boxes, recognizes the text

content using OCR, and classifies their role in the chart) and an idea that recovering a visual encoding specification based on both inferred text and chart type information. The pipeline is a novel work. The fundamental contribution of this paper is the continuation and improvement of ReVision.

- Advantages:

- The previous work cited in this paper is adequate.
- As shown in Section 6.2, based on the pipeline and the technique, recovering specification indeed realizes a high accuracy.
- Detailed information, including a summary of the various problem we might encounter with this technique, are introduced.
- It is good to see there is an open source code.
- The supplementary material is helpful to understand the process, especially the Appendix B.

- Disadvantages:

However, I wonder whether this pipeline has a strong connection to visualization. Recovering the visual encoding specification makes sense to charts redesign and visualization evaluation. But, this pipeline can be applied in the many fields like identifying the information on outdated magazines or newspapers. This pipeline lacks visualization-related features. Compared to EuroVis, I consider that this paper is more suitable for image identification.

Secondly, the specifications that can be recovered with this method are still limited. For example, error bars, box plots are common as well. Besides, if the author focuses on text contents, word cloud is a good experimental object.

Thirdly, I cannot find how to deal with the discontinuous axes.

Fourthly, as for figures, I think author should highlight the errors in Figure 8. In addition, I suggest that some statistical tables could be replaced by charts, which are more conducive to read.

Finally, I think it is necessary to add a section about summary and feature description of the diverse types of text appearing in the charts.

1.2 RelationLine

I had a discussion with Jing Xia about what needs to be improved and modified:

- Motivation need to be improved.
- “id” and “date” need to be hidden from the interface.
- We need a new color encoding scheme based on k-means to map the similar positions into similar colors.
- The ego network view need to be redesigned.
- A part of data processing steps is changed.
- Update related contents in paper.

1.3 New Idea

We collected relevant data and readied for brainstorming. In addition, I sent the power point to JiaKai.

2 To Do

2.1 RelationLine

Modify the code and paper.

2.2 Reading

About Deep Learning or Neural Network.

2.3 Final review